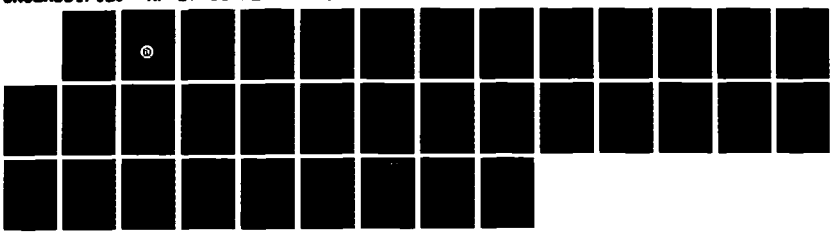
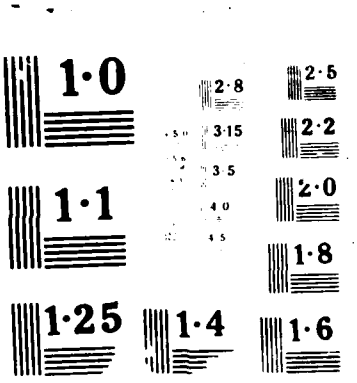
AD-A170 541 SELF-CRITICAL OR GENERALIZED LIKELIHOOD ANALYSIS OF VARIANCE WITH APPLICA..(U) RENSSELAER POLYTECHNIC INST TROY NY SCHOOL OF MANAGEMENT C E LAWRENCE ET AL. 1986 1/1

UNCLASSIFIED WP-37-86-P20 ARO-18072.16-MA F/G 12/1 NL




1·0
1·1
1·25
2·8
3·15
3·5
4·0
4·5
2·5
2·2
2·0
1·8
1·4
1·6

Rensselaer Polytechnic Institute
School of Management

Working Paper No. 37-86-P20

DISCLAIMER NOTICE

THIS DOCUMENT IS BEST QUALITY PRACTICABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

SELF-CRITICAL OR GENERALIZED LIKELIHOOD ANALYSIS OF VARIANCE WITH APPLICATION TO PROTHROMBIN TIMES

by

C. E. Lawrence,
A. S. Paulson, and
G.R. Swope

School of Management
Rensselaer Polytechnic Institute
Troy, NY 12180-3590
(518)266-6586

Self-Critical or Generalized Likelihood Analysis of Variance with Application to Prothrombin Times

by

C.E. Lawrence
New York State Division of Laboratories

A.S. Paulson*

and

G.R. Swope
Rensselaer Polytechnic Institute

* Research sponsored in part by U.S. Army Research Office Contract DAAG29-81-K-0110.

Self-Critical ANOVA with Application to Prothrombin Time

Summary

The purpose of this work was to develop procedures which would improve the uniformity of the grading procedure in the hematology proficiency testing program of the New York State Department of Health. Approximately 400 laboratories in New York participate in a program of clinical laboratory proficiency testing, with each laboratory using one of eight methods and one of nine thromboplastins. The responses of each testing program may be modeled as a two-way layout, some cells of which are empty. A model-critical analysis of variance technique was used to determine simultaneously, and in the presence of out-of-control laboratories i.e., outliers, the effects of method and thromboplastin on prothrombin time. The word model-critical indicates that, according to the model, set up initially on a tentative basis so as to allow for further evolution, the responses should have the structure of a two-way layout with interaction, and that the responses should have a common error distribution. This tentative framework is examined critically by varying the way obeservational information is processed to produce parametric summaries. If the summarizations are insensitive to the variation in information processing, then the tentative model stands; if not, the tentative model must be evolved. We provide an objective means of statistically assessing variations in summarization.

The model-critical analysis produced a common standard deviation, identified out-of-control laboratories and produced a narrower acceptable range of reported prothrombin times and thus improved the efficiency of the grading procedure. For proficiency testing no advantage was found in the use of either a common thromboplastin or freeze-dried, coumarinized patient plasmas rather than artifically depleted commerical plasmas, except for special purposes.

Keywords: proficiency testing; model-critical estimation; outliers, prothrombin time; generalized likelihood, information divergence, thromboplastins

1. INTRODUCTION

The New York State Department of Health operates a program of clinical laboratory proficiency testing. In this testing program, samples are sent periodically in mailout kits to all laboratories testing human specimens from the residents of the State. For many clinical sample types, it is not possible to prepare a sample in which the parameters are known in advance; they must be estimated from the data. The results of the test are graded on two bases. If any result contains an error of such a magnitude as to jeopardize the well-being of a patient, the result is unsatisfactory. In order to detect problems before they adversely affect a patient's health, statistical quality control concepts are also used as a basis for grading. Under this combination of concepts, one would hope to find that the bulk of the results would follow a well-behaved error model with a small number of outliers that represent the out-of-control laboratories.

In this report we will focus on the testing of proficiency of prothrombin time, a measure of plasma clotting time. Prothrombin time is used principally to monitor the status of patients being treated by oral anticoagulant drug therapy. An overdosage of an anticoagulant such as warfarin may lead to hemorrhagic complications, while underdosage increases the risk of thromboembolic complications. This test is also used routinely to screen patients to detect those at risk of bleeding excessively. For both purposes, the prothrombin time must be accurately standardized and correctly interpreted.

The effect of thromboplastin on prothrombin time has been studied by a number of investigators (Loeliger et al. (1984); Biggs and Denson (1967); Ingram and Hills (1976); Loeliger et al. (1976); Poller (1975)). Reference thromboplastins for calibration have recently become available (Hermans and von den Besselaar (1983);

von den Besselaar et al. (1984)) but these have been used infrequently in the United States. The simultaneous effects of method and thromboplastin have also been reported Evatt et al. (1981); Triplett et al. (1984); von den Besselaar et al. (1984)). Proficiency testing is complicated by the fact that prothrombin time measurements vary not only with the skill of the performer but also with the method and the thromboplastin, a clot initiation reagent, used. In proficiency testing, it is necessary to take into account structural effects of method and thrombolplastin as well as the error model while avoiding the excessive influence of the outlying, or out of control, observations. The model we make use of in order to account for all the inter-relationships is a two-way analysis of variance with interaction and a large proportion of empty cells.

In the next section, we will develop a model-critical or self-critical estimation procedure. The specifics of materials and data are given in later sections. The fourth section describes our application of the self-critical method to the prothrombin proficiency testing problem.

2. Model-Critical or Self-Critical Analysis of Variance

We provide first a generalization of the log likelihood for the normal distribution and then generalize this log likelihood to a two-way layout with interaction and empty cells. Let $f(x;\ \mu,\sigma^2)$ represent the Gaussian density with mean μ and variance σ^2. It is easy to show that

$$\int_{-\infty}^{\infty} f^{1+c}(x;\ \mu,\sigma^2)dx = Q(\mu,\sigma^2,c) = [(2\pi\sigma^2)^c(1+c)]^{-\frac{1}{2}}, \qquad (2.1)$$

$-1 < c < \infty$, is the information generating function (Golomb, 1966) of $f(x)$. For notational convenience we shall often delete the arguments of functions when

misinterpretation is not possible; for example, we will sometimes write $f(x)$ or simply f for $f(x;\ \mu,\sigma^2)$. Observe that $Q(\mu,\sigma^2,0) = 1$ and that the information-theoretic properties of $f(x)$ may be developed from Q, for example the entropy of $f(x;\ \mu,\sigma^2)$ is $-\partial Q(\mu,\sigma^2,0)/\partial c$. Our main objective is to develop a model-critical estimation procedure for μ and σ^2 of $f(x)$ based on the equation (2.1). This is accomplished by rearranging (2.1) to get

$$\int_{-\infty}^{\infty} \frac{f^{1+c}}{Q}\, dx = 1. \tag{2.2}$$

If we differentiate (2.2) with respect to $\theta = \mu$ and σ^2 we get

$$\int_{-\infty}^{\infty} f\left\{f^c\left[(1+c)\,\frac{\partial \log f}{\partial \theta} - \frac{\partial \log Q}{\partial \theta}\right]\right\}dx = 0. \tag{2.3}$$

The construction embodied in (2.2) and (2.3) shows that if x is a random variable with density f, then the expectation

$$E\left\{f^c\left[(1+c)\,\frac{\partial \log f}{\partial \theta} - \frac{\partial \log Q}{\partial \theta}\right]\right\} = 0. \tag{2.4}$$

Accordingly, if $x_1, x_2, \ldots, x_n$ is a random sample from $f(x;\ \mu,\sigma^2)$, then setting $\theta = \mu$ and $\theta = \sigma^2$ in

$$\sum_{i=1}^{n} f^c(x_i)\left[(1+c)\,\frac{\partial \log f(x_i)}{\partial \theta} - \frac{\partial \log Q}{\partial \theta}\right] = 0, \tag{2.5}$$

forms a set of estimating equations for $\theta = \mu$ and σ^2. The line of argument from (2.1) to (2.5) is parallel to that of maximum likelihood (see for example, Kendall and Stuart, Vol. II, pp. 8-10). Indeed, the equations (2.5) are those of the (log) likelihood estimators when $c=0$. The estimators for μ and σ^2 satisfy the implicit equations

$$\mu = \sum_{i=1}^{n} w_i(c)\, x_i,$$

$$\sigma^2 = (1+c) \sum_{i=1}^{n} w_i(c)\, (x_i-\mu)^2, \tag{2.6}$$

$$v_i(c) = f^c(x_i;\ \mu,\sigma^2), \quad v.(c) = \sum_{i=1}^{n} v_i(c).$$

$$w_i(c) = v_i(c)/v.(c)$$

The estimators for μ and σ^2, $\hat{\mu}(c)$ and $\hat{\sigma}^2(c)$ say, have been constructed so as to impose an adaptive "Gaussian screen" or "template" on the data. The action of this screen is accomplished through the factor $f^c(x)$ in (2.3) and through $f^c(x_i)$ in (2.5). The screen is adaptive because the values of μ and σ^2 are not known but in the process of iterating in (2.6) the estimators $\hat{\mu}(c)$ and $\hat{\sigma}^2(c)$ finally settle on values which are most consistent with the working model of Gaussianity and the data. If, for example, an outlying observation x_i not consistent with Gaussianity is present in the set $x_1, x_2, \ldots, x_n$, then the screen will impose the weight $f^c(x_i : \hat{\mu}(c), \hat{\sigma}^2(c))$ on the observation x_i so that its contribution to the estimates of μ and σ^2 will be small. This particular x_i will almost always require further study or attention but in some cases, as in this paper, these observations need to be removed from the original data set as well as investigated. The action of the screen is not limited to a single observation. We have found these methods to be useful in practice even when as many as 40% of the original sample were highlighted as not consistent with a single Gaussian population.

The estimating equation (2.5) may be regarded as a differential equation whose solution may be shown to be the objective function, the generalized likelihood,

$$\ell_c = \frac{1}{c} \sum_{i=1}^{n} \left\{ \frac{f^c(x_i : \mu, \sigma^2)}{Q^a(\mu, \sigma^2, c)} - 1 \right\}, \quad c \neq 0, \tag{2.7}$$

where $a = c/(1+c)$. The objective function ℓ_c is a generalization of the log likelihood; indeed the limit of ℓ_c as c tends to zero reduces to

$$\ell_0 = \sum_{i=1}^{n} \log f(x_i : \mu, \sigma^2), \tag{2.8}$$

the usual log likelihood.

The expression (2.7) is easily extended to cover a Gaussian error structure

combined with functional structure. Accordingly we shall develop an objective function for the two-way analysis of variance layout with interaction and empty cells. The variation of the user specified c parameter corresponds to variation in the way the information is extracted from the sample to arrive at model specification.

The family of estimators or model summarizations which result from variation of the user-specified coefficient c should not change much if the underlying tentative model and the data are internally consistent. Indeed, extensive simulation trials completely corroborate this statement; on the other hand, if model and data are not consistent, the estimates can vary dramatically. This is particularly true if the data contains outliers vis-a-vis the error distribution assumption. Because all of our analyses center on the processing of sample information to arrive at a summarization of a tentative model, we will call the procedure which enables this self- or model- critical. Models are evolved when this process of model-criticism (Box, 1979; Daniel 1978; Paulson and Nicklin, 1983; Paulson and Delehanty, 1983) shows that the data and the model, both error and function components, are not internally or mutually consistent.

As we shall see, an appropriate model for the hemotology testing program is

$$y_{ijk} = \log(x_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_{ij} + u_{ijk}, \qquad (2.9)$$

for $i = 1,2,..,\ I,\ j = 1,2,...,\ J,\ k = 1,2,...,\ n_{ij}$, and where u_{ijk} are independent normal (Gaussian) with mean zero and variance σ^2, in short u_{ijk} are $N(0,\sigma^2)$. The x_{ijk} represent the prothrombin time for the k^{th} laboratory using method i and thromboplastin j in our context. The generalized likelihood for the model (2.9) is

$$\ell_c = \frac{1}{c} \sum_i \sum_j \sum_k \left[\frac{f^c(u_{ijk})}{Q^\delta} - 1 \right], \qquad (2.10)$$

where

$$f(u_{ijk}) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\,(x_{ijk} - \mu - \alpha_i - \beta_j - \gamma_{ij})^2/\sigma^2\right\}$$

and Q is as before. On differentiating ℓ_c with respect to μ, α_i, β_j, γ_{ij}, and σ^2 and setting the resulting expressions to zero we obtain the system of implicit equations

$$\sum_i \sum_j \sum_k (y_{ijk} - \mu - \alpha_i - \beta_j - \gamma_{ij})\, v_{ijk}(c) = 0 \qquad (2.11a)$$

$$\sum_j \sum_k (y_{ijk} - \mu - \alpha_i - \beta_j - \gamma_{ij})\, v_{ijk}(c) = 0, \quad i=1,2,\ldots,I, \qquad (2.11.b)$$

$$\sum_i \sum_k (y_{ijk} - \mu - \alpha_i - \beta_j - \gamma_{ij})\, v_{ijk}(c) = 0, \quad j=1,2,\ldots,J, \qquad (2.11.c)$$

$$\sum_k (y_{ijk} - \mu - \alpha_i - \beta_j - \gamma_{ij})\, v_{ijk}(c) = 0, \quad \begin{matrix} i=1,2,\ldots,I \\ j=1,2,\ldots,J, \end{matrix} \qquad (2.11.d)$$

$$\sigma^2 = \frac{(1+c) \sum_i \sum_j \sum_k (y_{ijk} - \mu - \alpha_i - \beta_j - \gamma_{ij})^2\, v_{ijk}(c)}{\sum_i \sum_j \sum_k v_{ijk}(c)}, \qquad (2.11.e)$$

and

$$v_{ijk}(c) = \exp\left\{-\frac{c}{2}\,(y_{ijk} - \mu - \alpha_i - \beta_j - \gamma_{ij})^2/\sigma^2\right\}. \qquad (2.11.f)$$

The system (2.11) is not of full rank for $c \geq 0$. In order to obtain a full rank system we impose the constraints - and these are the most natural -

$$\sum_i \alpha_i v_{i..}(c) = 0, \quad \sum_j \beta_j v_{.j.}(c) = 0, \qquad (2.12.a)$$

$$\sum_i \gamma_{ij} v_{ij.}(c) = 0, \quad j=1,2,\ldots,J, \qquad (2.12.b)$$

$$\sum_j \gamma_{ij} v_{ij.}(c) = 0, \quad i=1,2,\ldots,I, \qquad (2.12.c)$$

and where a dot indicates summation over a subscript, e.g.,

$$v_{ij.}(c) = \sum_k v_{ijk}(c), \quad v_{i..}(c) = \sum_j \sum_k v_{ijk}(c).$$

This system of equations is solved by recursion. Programs are available on request from the authors.

Equations (2.9)-(2.12) determine the generalized likelihood or model-critical analysis for a two-way layout with ($n_{ij} = 0$ for some i and j) or without ($n_{ij} > 0$) empty cells. When cells of the array are empty, the same equations are applicable but with some minor modifications to be discussed in section 4 in the context of the analysis of the prothrombin time data.

Typical values of c to be used in the model-critical procedure involving a two-way layout with a normal error distribution are $0 \leq c \leq 0.50$. As c varies from c = 0 to c = .1, .2, .3, for example, the parameter estimates surface $\hat{\theta}(c) = (\hat{\mu}(c), \hat{\alpha}_i(c), \hat{\beta}_j(c), \hat{\gamma}_{ij}(c), \hat{\sigma}^2(c), i = 1,2,\ldots, I, j = 1,2,\ldots,J)$, say, will all vary as a function of c. We need to be able to objectively assess whether the response surface varies statistically significantly as a function of c. The next section provides procedures for making this assessment.

3. Tests of Fit

First we consider the case when $x_1, x_2, \ldots, x_n$ is tentatively taken to be independent $N(\mu,\sigma^2)$. The process of varying the way the information in the data is summarized in the parameters μ and σ^2 through the generalized likelihood will lead to families of summarizations $\hat{\mu}(c)$ and $\hat{\sigma}^2(c)$. For two different values of c, c and c', say, the values of $\hat{\mu}(c)$ and $\hat{\mu}(c')$, $\hat{\sigma}^2(c)$ and $\hat{\sigma}^2(c')$ can be different. These differences can be used to develop a statisical test of fit for the appropriateness of the model. Given the estimates $\hat{\mu}(c)$ and $\hat{\sigma}^2(c)$, an estimate of the model density is

$$f(x;\ \hat{\mu}(c),\hat{\sigma}^2(c)) = (2\pi\hat{\sigma}^2(c))^{-\frac{1}{2}} \exp(-\tfrac{1}{2}(x-\hat{\mu}(c))^2\hat{\sigma}^2(c)). \qquad (3.1)$$

This estimate of the model density captures the sample information provided the tentative Gaussian model is correct. A test of appropriateness of the Gaussian model for the data $x_1, x_2, \ldots, x_n$ can be based on

$$D(1:2:c) = n \int_{-\infty}^{\infty} (f(x: \hat{\mu}, \hat{\sigma}^2) - f(x: \hat{\mu}(c), \hat{\sigma}^2(c)) \log \frac{f(x: \hat{\mu}, \hat{\sigma}^2)}{f(x: \hat{\mu}(c), \hat{\sigma}^2(c))} dx \quad (3.2)$$

for some $c \neq 0$, where $D(1:2:c)$ represents an information divergence (Kullback, 1959, Chapters 1 and 2) based on the assumed model and the data as summarized in the estimators $\hat{\mu}(c)$ and $\hat{\sigma}^2(c)$, $c \geq 0$. We have extensively investigated this statistic as a test of fit for a variety of values of c and found that it does indeed make for a good test of Gaussianity. However, the percentage points of this statistic are appropriate for the case of testing that $x_1, x_2, \ldots, x_n$ are independent Gaussian $N(\mu, \sigma^2)$ but would not be appropriate for testing a model with combined Gaussian error and functional structure since the required percentage points depend on the specifics of the functional model structure. Thus D could not be used to test the appropriateness of (2.9) for prothrombin times.

However, another information divergence which explicitly depends only on estimated variances $\hat{\sigma}^2(c)$, and which does not depend on the specifics of the functional structure of the model is

$$J(1:2:c) = n \int_{-\infty}^{\infty} (f(x: 0, \hat{\sigma}^2) - f(x: 0, \hat{\sigma}^2(c)) \log \frac{f(x: 0, \hat{\sigma}^2)}{f(x: 0, \hat{\sigma}^2(c))} dx \quad (3.3)$$

$$= \frac{n}{2} \left[\frac{\hat{\sigma}^2}{\hat{\sigma}^2(c)} + \frac{\hat{\sigma}^2(c)}{\hat{\sigma}^2} - 2 \right].$$

The statistic J represents the divergence between an estimate of a Guassian density based on the maximum likelihood estimate of the variance and an estimate of the Gaussian density based on the generalized likelihood estimate of variance. In this case any common mean substituted for zero in (3.3) will yield the same result. This

does not imply that (3.3) is independent of the estimates of the mean as may be seen from system (2.6).

The statistic (3.3) can be viewed as an analogue of the Shapiro-Wilk W statistic for testing for Gaussianity (Shapiro and Wilk, 1965). The ratinale behind

$$W = \left[\sum_{j=1}^{n} a_j x_j\right]^2 / \sum_{j=1}^{n} (x_j - \hat{\mu}^2)^2, \tag{3.4}$$

where $x_1, x_2, \ldots, x_n$ is putatively $N(\mu,\sigma^2)$ and the a_j are tabulated constants (Shapiro, 1980), is that the numerator and denominator of (3.4) are both estimates of a constant multiple of σ^2. If one of the estimators is markedly different from the other, small values of W will result and evidence against the hypothesis of Gaussianity will be strong.

The rationale behind the nonnegative statistic $J(1:2:c)$ of (3.3) is similar and is as follows. If $x_1, x_2, \ldots, x_n$ are independent $N(\mu,\sigma^2)$, then, apart from sampling error, both $\hat{\sigma}^2$ and $\hat{\sigma}^2(c)$ are estimators for σ^2. Large values of $J(1:2:c)$ will provide evidence against the hypothesis of Gaussianity. The statistic $J(1:2:c)$ is particularly sensitive to outlier-like departures from Gaussianity since the influence curves (see Barnett and Lewis, 1978, pp. 136-142) at observation x_j for the estimators of μ and σ^2 at the univariate Gaussian density are proportional to

$$f^c(x_j;\mu,\sigma^2)\left\{(1+c)\,\frac{\partial \log\, f(x_j:\mu,\sigma^2)}{\partial\theta} \;-\; \frac{\partial \log Q}{\partial\theta}\right\}, \quad c > 0,$$

for $\theta = \mu$ and σ^2 respectively. Therefore, in the process of the generalized likelihood's adaptation to the best summarization of the data consistent with the tentative model of Gaussianity, the influence of an outlying observation or groups of outlying observations will be ultimately downweighted by $f^c(x_j: \hat{\mu}(c),\hat{\sigma}^2(c))$. For x_j far removed from $\hat{\mu}(c)$ in comparison with the scale $\hat{\sigma}(c)$, $f^c(x_j: \hat{\mu}(c),\hat{\sigma}^2(c))$ will be nearly zero.

Table 1 provides the upper percentage points for the statistic $J(1:2:c)$ for $c = -.2, -.1, .1, .2, .3, .4, .5$ and for $n = 10, 20, 24, 30, 40, 60, 120, 480$. These percentage points have been developed as follows. First, 10,000 independent realizations of $J(1:2:c)$ were simulated for each value of c and n. Next these realizations were put in ascending order and appropriate estimates of the percentage points were tabulated. Finally, cubic spline functions were fit to the resulting surfaces in sample size n, coefficient c, and size of test α. Completely independent simulations, i.e., independent programmers and programs, were used to check the simulations which produced Table 1.

In order to determine the dependence of the percentage points J of Table 1 on the specifics of a functional model sturcture, the percentage points of J were simulated under a variety of regression structures, linear model structures, and nonlinear functional model structures, all under the assumption of an additive Gaussian error structure. In all cases the percentage points of the statistic J were the same as those computed under the model assumption that a Gaussian error structure alone describes the data, i.e., $x_1, x_2, \ldots, x_n$ are independent $N(\mu, \sigma^2)$, apart from sampling error and a correction which accounts for the number of parameters estimated. This independence of J of the functional structure is suggested by the similar property of the Shapiro-Wilk statistic but we have not succeeded in developing an analytical proof.

Consider the modeling framework where it is postulated the responses of interest follow the tentative functional and error representation

$$y_i = h(x_{1i}, x_{2i}, \ldots, x_{mi};\ \theta_1, \theta_2, \ldots, \theta_q) + e_i,\ i = 1, 2, \ldots, n,$$

where h is some specified function which may simply be constant μ, the additive errors e_i are independently and identically $N(0, \sigma^2)$, the x_{ji} are measurements on variables which may influence the y_i, and $\theta_1, \theta_2, \ldots, \theta_q$ are parameters to be estimated from

the data. The tentative model will be fit by generalized likelihood for c=0 and some c>0, say c=.3, with resulting estimates $\dot{\theta}_1(c)$ and $\dot{\sigma}^2(c)$. The statistic J(1:2:c) is this setting computed as

$$J(1:2:c) = \frac{1}{2}(n-q+1)\left[\frac{\dot{\sigma}^2}{\dot{\sigma}^2(c)} + \frac{\dot{\sigma}^2(c)}{\dot{\sigma}^2} - 2\right]$$

and referred to Table 1. If J(1:2:c) exceeds the tabulated critical value at level α, then it is possible to improve the tentative model in the sense that some data may be inconsistent with the model, the functional structure of the model may need to be evolved, the error structure of the model may need to be evolved, or both error structure and functional structure should be examined and evolved, etc. In some cases considerable study may be required before the source of a statistically significant J(1:2:c) is found and a choice for evolution of the model is made, although in many cases, such as here for prothrombin times, the source of a statistically significant J(1:2:c) may be readily found. Standard graphical and diagnostic procedure should be used in conjunction with the generalized likelihood, and the generalized likelihood residuals and the J test. Indeed, we regard all aspects of the generalized likelihood as complementary to the standard procedures of data analysis and statistics. We note in passing that the statistic J of (3.3) is readily extended to the p-variate case.

Even though the statistic J(1:2:c) has been developed to aid in model assessment and evolution, it provides for a good pure test for normality. Table 2 presents powers of the J, skewness b_1, kurtosis b_2, the Anderson-Darling, and the Shapiro-Wilk tests for normality under the alternatives of a heavily right-skewed lognormal distribution, t distributions on m degrees of freedom (designated T(m)), chi-squared distributions m m degrees of freedom (designated $\chi^2(m)$), and four mixture alternatives, Mx1-Mx4. The mixture alternatives are as follows: Mx1 is 75%

N(0,1) and 25% N(2,1); Mx2 is 50% N(0,1) and 50% N(2,1); Mx3 is 50% N(0,1) and 50% N(0,4); Mx4 is 50% N(0,1) and 50% N(1,1). Mx4 is a particularly severe test of any test for normality since the existence of the mixture can be very difficult to detect. We have provided tabulations in Table 1 and Table 2 of values c = -0.2 and -0.1 because the use of negative c in the generalized likelihood has been found to be useful in several applications involving in-lying contamination of data and multiple clusters of data. Tabulations of powers in Table 2 indicates that J(1:2:c) provides for a good test of normality for a wide range of c. The importance of J(1:2:c), however, is due to its linking estimation and assessments of fit.

Example. In a study concerning tests for outliers, Quesenberry and David (1961) provide the sixteen observations .32,.35,.37,.38,.39,.44,.45,.46,.47,.48,.52,.53, .57,.74,.74, 1.09 in illustration of a studentized range test. This studentized range test finds the observation 1.09 to be too large, but just barely, to be consistent with the $x_1, x_2, \ldots, x_{16}$ being independent and identically Gaussian. We find $\hat{\sigma}^2 = \hat{\sigma}^2(0)$ = 0.0357, $\hat{\sigma}^2(.3)$ = 0.0134 and J(1:2:c) = 8.32. Comparison of this value of J with the critical values of Table 1 shows that this data is dramatically non-normal, a finding visually corroborated by a normal probability plot. The studentized range test experiences difficulty in rejecting the observation 1.09 as an outlier because the outlying nature of this data point is being inter-mixed with the otherwise dramatic non-normality of the data and the value 1.09 is also dramatically influencing the estimated variance of the sample.

The J(1:2:c) test statistic is useful in evaluating the statistical status of a tentative modeling structure and in determining whether a tentative model should be evolved. Furthermore, the use of diagnostic tools stemming from the generalized likelihood procedure and other diagnostic tools will usually be helpful in determining the direction of the evolution.

Table 1. Percentage Points of the Test Statistic $J(1:2:c)$ for $-0.2 \leq c \leq 0.5$

		Size of Test						
	Sample Size	0.75	0.80	0.85	0.90	0.95	0.975	0.99
	10	0.38	0.45	0.60	1.61	6.40	15.7	40.5
	20	0.46	0.56	0.70	1.11	2.70	5.32	10.8
	24	0.48	0.58	0.74	1.10	2.34	4.40	8.37
c=0.5	30	0.49	0.60	0.75	1.07	2.16	3.91	6.95
	40	0.51	0.63	0.80	1.06	1.95	3.31	5.60
	60	0.53	0.65	0.80	1.05	1.68	2.78	4.50
	120	0.54	0.67	0.84	1.10	1.66	2.33	3.64
	480	0.55	0.69	0.87	1.10	1.65	2.10	3.11
	10	0.24	0.28	0.34	0.61	2.45	6.92	19.7
	20	0.30	0.36	0.45	0.64	1.60	3.02	6.33
	24	0.31	0.38	0.48	0.66	1.45	2.64	5.07
c=0.4	30	0.32	0.40	0.50	0.68	1.35	2.41	4.40
	40	0.34	0.42	0.53	0.70	1.30	2.20	3.73
	60	0.36	0.44	0.55	0.72	1.12	1.85	3.15
	120	0.37	0.46	0.57	0.75	1.11	1.65	2.51
	480	0.37	0.47	0.60	0.85	1.09	1.51	2.10
	10	0.13	0.15	0.18	0.23	0.74	2.11	8.16
	20	0.17	0.21	0.28	0.34	0.77	1.46	3.19
	24	0.18	0.22	0.27	0.36	0.77	1.41	2.87
c=0.3	30	0.19	0.23	0.28	0.38	0.77	1.39	2.47
	40	0.20	0.24	0.30	0.42	0.74	1.31	2.15
	60	0.21	0.26	0.32	0.44	0.69	1.10	1.82
	120	0.22	0.28	0.34	0.45	0.67	0.96	1.49
	480	0.22	0.28	0.35	0.46	0.67	0.88	1.26
	10	0.059	0.068	0.078	0.095	0.18	0.48	1.33
	20	0.078	0.092	0.11	0.14	0.27	0.54	1.29
	24	0.082	0.097	0.12	0.15	0.29	0.56	1.23
c=0.2	30	0.087	0.10	0.13	0.16	0.31	0.57	1.12
	40	0.092	0.11	0.14	0.18	0.31	0.58	1.02
	60	0.099	0.12	0.15	0.20	0.32	0.53	0.92
	120	0.11	0.13	0.16	0.21	0.32	0.47	0.75
	480	0.11	0.14	0.17	0.23	0.33	0.46	0.69
	10	0.015	0.017	0.020	0.023	0.031	0.052	0.11
	20	0.020	0.023	0.028	0.034	0.051	0.095	0.23
	24	0.021	0.025	0.030	0.037	0.056	0.11	0.25
c=0.1	30	0.023	0.027	0.032	0.040	0.066	0.12	0.26
	40	0.024	0.029	0.035	0.045	0.072	0.13	0.25
	60	0.027	0.032	0.039	0.050	0.074	0.13	0.23
	120	0.029	0.036	0.044	0.057	0.085	0.12	0.21
	480	0.030	0.039	0.047	0.060	0.089	0.12	0.20

Table 1. Percentage Points of the Test Statistic J(1:2:c) for $-0.2 \leq c \leq 0.5$ (cont'd)

	Sample Size	0.75	0.80	0.85	0.90	0.95	0.975	0.99
	10	0.017	0.019	0.022	0.025	0.029	0.034	0.038
	20	0.022	0.025	0.029	0.035	0.044	0.055	0.093
	24	0.023	0.027	0.031	0.038	0.049	0.063	0.11
c=-0.1	30	0.025	0.030	0.035	0.042	0.054	0.071	0.13
	40	0.027	0.032	0.038	0.047	0.063	0.091	0.17
	60	0.030	0.036	0.043	0.054	0.075	0.11	0.20
	120	0.034	0.041	0.052	0.065	0.094	0.14	0.24
	480	0.038	0.047	0.060	0.075	0.11	0.16	0.28
	10	0.077	0.086	0.097	0.11	0.13	0.15	0.16
	20	0.097	0.11	0.13	0.15	0.18	0.22	0.28
	24	0.10	0.12	0.14	0.16	0.20	0.25	0.33
c=-0.2	30	0.11	0.13	0.15	0.18	0.22	0.27	0.42
	40	0.12	0.14	0.16	0.20	0.26	0.33	0.55
	60	0.13	0.16	0.19	0.23	0.31	0.40	0.70
	120	0.15	0.18	0.22	0.28	0.40	0.55	0.97
	480	0.17	0.21	0.26	0.33	0.49	0.70	1.23

Table 2. Comparative Powers of the J, b_1 and b_2, the Anderson-Darling (A-D), and Shapiro-Wilk (S-W) Statistics for Several Alternatives

(a) Size α = 0.1 and n = 20

	0.5	0.4	0.3	0.2	0.1	-0.1	-0.2	b1	b2	A-D	S-W
LN	0.84	0.82	0.80	0.77	0.72	0.59	0.52	--	--	0.94	0.96
T(9)	0.22	0.22	0.21	0.20	0.19	0.15	0.12	--	--	0.17	0.15
T(7)	0.27	0.26	0.25	0.24	0.22	0.17	0.15	--	--	0.19	0.19
T(5)	0.34	0.34	0.34	0.32	0.29	0.23	0.19	--	--	0.24	0.24
T(3)	0.52	0.52	0.51	0.48	0.46	0.37	0.32	--	--	0.41	0.42
T(1)	0.93	0.92	0.92	0.90	0.89	0.82	0.78	--	--	0.90	0.88
$X^2(14)$	0.20	0.20	0.20	0.19	0.18	0.16	0.15	--	--	0.24	0.28
$X^2(10)$	0.25	0.25	0.25	0.24	0.23	0.19	0.16	--	--	0.31	0.34
$X^2(6)$	0.32	0.33	0.32	0.31	0.28	0.22	0.20	--	--	0.43	0.48
$X^2(4)$	0.42	0.42	0.40	0.39	0.36	0.27	0.24	--	--	0.58	0.64
$X^2(2)$	0.63	0.62	0.58	0.55	0.51	0.39	0.34	--	--	0.86	0.91
$X^2(1)$	0.84	0.82	0.79	0.75	0.69	0.54	0.45	--	--	0.99	1.00
Mx1	0.10	0.09	0.08	0.08	0.08	0.08	0.09	--	--	0.16	0.15
Mx2	0.02	0.03	0.05	0.08	0.12	0.18	0.20	--	--	0.10	0.09
Mx3	0.31	0.31	0.30	0.27	0.23	0.15	0.11	--	--	0.22	0.21
Mx4	0.09	0.08	0.08	0.09	0.10	0.10	0.10	--	--	0.10	0.10

(b) Size α = 0.1 and n = 50

LN	0.99	0.99	0.99	0.98	0.97	0.92	0.87	1.0	0.94	1.0	1.0
T(9)	0.31	0.31	0.30	0.30	0.29	0.24	0.22	0.23	0.24	0.19	0.16
T(7)	0.40	0.41	0.40	0.40	0.38	0.31	0.28	0.29	0.32	0.26	0.23
T(5)	0.55	0.56	0.55	0.54	0.52	0.47	0.41	0.41	0.47	0.40	0.35
T(3)	0.80	0.80	0.79	0.78	0.76	0.71	0.64	0.58	0.72	0.69	0.61
T(1)	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.92	1.0	1.0	1.0
$X^2(14)$	0.30	0.31	0.31	0.30	0.29	0.25	0.23	0.59	0.25	0.50	0.56
$X^2(10)$	0.37	0.38	0.37	0.36	0.35	0.29	0.27	0.69	0.31	0.59	0.69
$X^2(6)$	0.50	0.50	0.50	0.48	0.45	0.38	0.33	0.86	0.40	0.82	0.89
$X^2(4)$	0.65	0.65	0.64	0.61	0.57	0.47	0.42	0.95	0.50	0.94	0.98
$X^2(2)$	0.88	0.87	0.85	0.83	0.79	0.68	0.61	0.99	0.72	1.0	1.0
$X^2(1)$	0.99	0.98	0.98	0.97	0.95	0.87	0.81	1.0	0.91	1.0	1.0
Mx1	0.10	0.09	0.07	0.07	0.08	0.08	0.09	0.22	0.08	0.25	0.25
Mx2	0.12	0.13	0.14	0.15	0.18	0.24	0.27	0.03	0.25	0.14	0.19
Mx3	0.46	0.46	0.44	0.41	0.38	0.28	0.24	0.26	0.31	0.30	0.21
Mx4	0.10	0.10	0.10	0.10	0.10	0.11	0.11	0.08	0.11	0.11	0.11

Table 2. Comparative Powers of the J, b_1 and b_2, the Anderson-Darling (A-D), and Shapiro-Wilk (S-W) Statistics for Several Alternatives (cont'd)

(c) Size $\alpha = 0.1$ and $n = 100$

	0.5	0.4	0.3	0.2	0.1	-0.1	-0.2	b1	b2	A-D	S-W
LN	1.00	1.00	1.00	1.00	1.00	0.99	0.98	1.00	1.00	1.00	1.00
T(9)	0.41	0.42	0.42	0.43	0.42	0.38	0.35	0.28	0.37	0.29	0.43
T(7)	0.56	0.57	0.57	0.57	0.55	0.51	0.46	0.35	0.50	0.36	0.54
T(5)	0.76	0.77	0.77	0.76	0.74	0.70	0.65	0.51	0.70	0.61	0.74
T(3)	0.95	0.95	0.95	0.95	0.94	0.92	0.88	0.71	0.93	0.89	0.94
T(1)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.96	1.00	1.00	1.00
$X^2(14)$	0.36	0.38	0.39	0.40	0.38	0.34	0.31	0.86	0.34	0.75	0.85
$X^2(10)$	0.48	0.49	0.50	0.50	0.49	0.43	0.38	0.94	0.44	0.86	0.94
$X^2(6)$	0.67	0.68	0.68	0.65	0.62	0.55	0.49	0.99	0.57	0.98	1.00
$X^2(4)$	0.84	0.84	0.83	0.82	0.79	0.71	0.64	1.00	0.73	1.00	1.00
$X^2(2)$	0.98	0.98	0.98	0.97	0.95	0.91	0.86	1.00	0.93	1.00	1.00
$X^2(1)$	1.00	1.00	1.00	1.00	1.00	0.99	0.97	1.00	0.99	1.00	1.00
Mx1	0.10	0.09	0.08	0.08	0.08	0.08	0.10	0.36	0.09	0.40	0.36
Mx2	0.28	0.29	0.29	0.30	0.31	0.36	0.39	0.02	0.39	0.22	0.15
Mx3	0.66	0.65	0.63	0.60	0.57	0.48	0.42	0.30	0.49	0.47	0.56
Mx4	0.09	0.09	0.09	0.10	0.10	0.11	0.12	0.08	0.11	0.10	0.11

4. Results of the Prothrombin Testing Program

Approximately 400 laboratories participate in the New York Department of Health prothrombin time proficiency testing program. This paper will focus on the approximately 320 of these 400 laboratories which use automated testing methods. The data analyzed in this study were reported for 12 test specimens, 3 in each of 4 specimen mailouts: January, July, and October of 1982 and July 1983. The laboratories which utilized automated testing methods used various combinations of eight analysis methods and nine thromboplastins. For example, for the July 1983 mailout, 318 automated laboratories used 47 combinations of methods and thromboplastins as indicated in Table 3. Twenty-five combinations of method and thromboplastin were not utilized by any laboratory.

Table 3 suggests that an appropriate model for the analysis of the clotting times is a two-way analysis of variance layout with interaction and with an error distribution to be determined. The development of a tentative error model for the responses from the participating laboratories was facilitated by a special identification study. In this special study, 13 reference laboratories (those with a reputation for excellence) were sent three specimens. Each reference laboratory was instructed to report triplicate measurements on the specimens obtained with each of the 9 types of thromboplastins supplied in the mailout kits. Analysis of the special study data showed that a two way layout provided a reasonable tentative model and that the logarithm of prothrombin times provided normal and homoskedastic residuals.

Therefore the tentative model assumed is that the participating laboratories will produce prothrombin times x_{ijk} which follow the model (2.9) where x_{ijk} is the prothrombin time for method i, thromboplastin j, i=1,2,..., 8, j=1,2,..., 9, and k=1,2,..., $n_{ij} \geq 0$. The quantity μ is the overall c grand mean, α_i is the effect due to method i, β_j is the effect

Table 3

Number of Laboratories Using Each Combination of Automated Method and Thromboplastin, July 1983

Method	Thromboplastin								
	A	B	C	D	E	F	G	H	I
1		2					1		
2		1		1	1	5	2		
3	5	40	4	2	7	59	19	1	2
4	1	12		1	2	8	8	2	13
5	1	5		1		3	5	3	10
6		2				1	2		
7		5				2	2		2
8	2	54	1	1	1	9	3	2	2
	9	121	5	6	11	87	42	8	29

Method Codes: 1, Autofi (Date); 2, Clotek (Hyland); 3, Fibrometer (BioQuest); 4, Coag-A-Mate 150 or Dual Beam (General Diagnostic); 5, Coag-A-Mate 2001 or X2 (General Diagnostics); 6, Coagulation Profiler (Bio Data); 7, Coagulizer (Sherwood); 8, Electra (Medical Laboratory Automation).

Thromboplastin Codes: A, Dade Activated Liquid; B, Dade C; C, Dade Reagent; D, Hyland Dried; E. Hyland Liquid; F, Ortho; G, Simplastin (General Diagnostics); H, Simplastin A; I, Simplastin Automated.

due to the combination of method i and thromboplastin j. The u_{ijk} are tentatively normal with mean zero and standard deviation σ.

The generalized likelihood for model (2.9) with all $n_{ij} > 0$ is given in (2.10). Table 3, however, shows that a number of the n_{ij} are 0, that is cells (1,A), (1,C), (1,F),... are empty so that the generalized likelihood is not directly applicable. We overcome this difficulty for c>0 as follows. To cell (1,A) we assign a large artificial value K, K>0; to cell (1,c) we assign the artificial value -K; to cell (1,K) we assign the value +2K; to cell (1,H) we assign the value -2K;.... The process continues until all empty cells have been assigned an artificial value. A suitable magnitude for K is determined by the location and spread of the basis data. In our case all the logarithms of prothrombin times are of the order 3 with a standard deviation of order .1 so that a value of K=10 would suffice. The strategy of assigning isolated artificial values +K, -K, +2K, ... is required in order to preclude the formation of a cluster of artificial outliers. In the maximization of (2.10) modified by inclusion of artificial outliers, the influence these artificial outliers exert on the estimates of μ, α_i, β_j, γ_{ij} and σ^2 will be at most of order

$$\exp\left[-\frac{c}{2}\left[\frac{K}{\sigma}\right]^2\right],$$

a quantity which is by construction virtually zero. This procedure for missing values requires that every row and every column contain at least one non-empty cell.

The model-critical or self-critical two-way analysis of variance procedure was applied to the twelve mailouts mentioned above. We provide only representative summaries of these analyses. Table 4 provides the results for grand means μ and standard deviations σ of the analyses for c = .1, .2, .25, and .3 and for the missing value analysis of variance (ANOVA). In every case, the grand means remain virtually the same across the usual least squares ANOVA and generalized likelihood

analysis with c = .1, .2, .25, and .3. Also in every case, the estimate of the standard deviation decreases dramatically with an increase in c to 0.1 and decreases little thereafter.

For c = .3 and sample number 1, for example, we compute

$$J(1;2;c) = \frac{246}{2}\left[\left[\frac{.049}{.071}\right]^2 + \left[\frac{.071}{.049}\right]^2 - 2\right]$$

$$= 70.8,$$

which is strongly statistically significant according to Table 1 and therefore indicates that the model and data are not consistent as they stand. The multiplier 246 is n-(I+J-1)-(I-1)(I-1) where n = 318, the number of participating laboratories, I = 8, the number of methods, and J = 9, the number of thromboplastins. The decrease in estimated standard deviation is due to the removal of the influence of the out of control laboratories, whose prothrombin times are not consistent with the two-way layout with interaction and with a normal error distribution. These out of control laboratories are easily determined by an examination of the generalized likelihood residuals.

Similar results concerning standard deviation are obtained for all twelve mailouts. This behavior is of course related to the estimates of μ, and α_i, β_j, and γ_{ij} for (missing value) maximum likelihood or (missing value) least squares and for the generalized likelihood. Sample 12, the last sample in the mailout of July 1983 provides a typical illustration of the behavior of the estimates of the parameters α_i and β_j as we move from least squares to generalized likelihood (c = 0.25). These results are presented in Table 5. Similar results obtain for the estimates of the γ_{ij}. These variations assume their true importance when the estimates are all combined to obtain the predicted cell means. Once the cell means and the estimated standard deviation are available, standardized grades for each laboratory are obtained. A

standardized grade is the number of standard deviations by which a particular laboratory differs from the predicted cell mean. For example, if a laboratory's prothrombin time results in a standardized grade of +3, this indicates that the laboratory's reported time was 3 estimated σ units above the cell (i.e., method and thromboplastin combination) mean.

The selection of a critical value beyond which a laboratory's performance would be considered out of control, required the responsible decision makers to trade off two types of error. Dollar costs or another single measurement unit for the relative weight to be assigned for falsely failing a satisfactory lab versus falsely passing an unsatisfactory lab were and are not available. Thus, the selection of a critical value was based on the judgment of the program administrators after a review of the implications of the various alternatives. Program administrators settled on a type I error level of 2%. Thus, the critical interval was set at ± 2.3 standard units about each cell mean. This represented a substantial reduction in the size of the interval from the three standard units that had been used with the prior grading system.

The practical effect of the application of self-critical estimation on the grading can be illustrated by considering the differences in the laboratories falling outside the critical region. In Table 6 we give a list of the laboratories falling outside the critical region using least squares estimation as well as those determined from model-critical estimation with $c=.25$.

Table 4

Estimates of generalized likelihood grand means and standard deviation σ of model (2.9) for 12 proficiency study mailouts.

	Sample	Missing Value ANOVA	Model-Critical Estimates c=0.10	c=0.20	c=0.25	c=0.30
	January 1982					
	Grand mean	2.44	2.44	2.44	2.44	2.44
1	SD	.071	.053	.050	.050	.049
	Grand mean	2.87	2.86	2.86	2.86	2.86
2	SD	.071	.051	.046	.044	.043
	Grand mean	3.05	3.04	3.04	3.04	3.04
3	SD	.077	.054	.050	.048	.047
	July 1982					
	Grand mean	2.54	2.54	2.54	2.54	2.54
4	SD	.056	.056	.054	.054	.053
	Grand mean	2.76	2.75	2.75	2.75	2.75
5	SD	.071	.059	.058	.057	.057
	Grand mean	2.99	2.99	2.99	2.99	2.99
6	SD	.063	.059	.058	.057	.056
	October 1982					
	Grand mean	2.51	2.51	2.51	2.50	2.50
7	SD	.063	.050	.048	.047	.046
	Grand mean	2.99	2.99	2.99	2.99	2.99
8	SD	.071	.057	.055	.054	.052
	Grand mean	3.21	3.21	3.21	3.20	3.21
9	SD	.077	.074	.071	.070	.069
	July 1983					
	Grand mean	2.49	2.49	2.49	2.49	2.49
10	SD	.063	.051	.050	.049	.048
	Grand mean	2.89	2.89	2.89	2.89	2.89
11	SD	.067	.056	.054	.054	.053
	Grand mean	3.12	3.12	3.12	3.12	3.12
12	SD	.077	.060	.058	.057	.056

Table 5: Out of Control Laboratories:
Cell Means and Standardized Grades

Cell	Least Squares $\hat{\sigma}$ = .099 Cell Mean	Least Squares Standardized Grades	Model-Critical (c=.25) $\hat{\sigma}(.25)$ = .057 Cell Mean	Model-Critical (c=.25) Standardized Grades
(3,F)	3.171	-5.57	3.169	-7.51
		-2.73		-3.67
		2.28		3.11
		-1.77		-2.35
(8, B)	3.051	-7.44	3.077	-10.54
		-1.94		-3.09
		2.23		2.55
		2.80		3.44
(3, A)	3.047	1.32	3.149	2.40
(3, 6)	3.147	2.03	3.303	2.92
(5, A)	3.044	2.68	3.250	4.52
(1, B)	3.210	1.31	3.109	-2.42

Note that only 5 of 318 laboratories are judged as being out of control using the least squares estimates (6.38 were expected) while with the model-critical procedure 12 of 318 observations fall in the critical region. This is the case since first, under least squares estimation with missing values, the outlying observations have the greatest influence on the parameter estimates and second, the iterative process involving the assignment of expected values to empty cells imparts an influence to the estimate of the standard deviation. The self-critical procedure imparts by construction zero influence to the estimate of the standard deviation. When the out-of-control laboratories are removed from the data base and the data re-analyzed in accordance with the two-way layout with normal error and interaction model, the parameter estimates, including $\hat{\sigma}(c)$, are virtually unchanged for c = .1, .2, .25, and .3. The re-computed value of J(1:2:c) is no longer significant for any value of c in $-0.2 \leq c \leq 0.5$.

The analysis and results described for mailout 12 of the New York State Department of Health hematology proficiency testing program are typical of those obtained for the remaining 11 studies. In each case the out-of-control laboratories are quickly and easily identified with resultant high-quality service supplied by the laboratories.

5. Conclusions

The grading system for this proficiency testing program is aimed at detecting laboratories whose results depart from the structural and error model that governs the majority of the testing laboratories. Even after the form of the appropriate error model had been identified, parameter estimators robust against the high influence of outlying observations are essential. Self- or model critical estimation which allows

for the simultaneous robust estimation of both location and scale parameters in the presence of empty cells is thus well suited to this problem.

This approach substantially improved the efficiency of the proficiency testing grading procedure in New York State as evidenced by the reduction averaging 20% in the standard deviation estimates, the almost routine detection of out-of-control laboratories, and the evolution to a much more uniform grading procedure.

The model-critical procedure incorporated the tentative model in making an assessment of the internal consistency of the model and the data through the parametric modification of the way information concerning a tentative model is extracted from the data. In the hemotology testing program some data was found to be inconsistent with the model but this need not always be the case; it is often the case that the proposed model is inadequate to deal with the data.

ACKNOWLEDGEMENTS

We acknowledge the advice on the software assistance provided by Dr. Ho-Ling Hwang and Mr. Ivan Auger. We thank Drs. W. Lawson and J. Dodds for their continuing interest and helpful discussions. The support of the work of A.S. Paulson by the Army Research Office under contract DAAG29-81-K-0110 is gratefully acknowledged.

REFERENCES

1. Barnett, V. and Lewis, T. (1978). Outliers in Statistical Data. New York: Wiley.

2. Biggs, R., Denson, K.W.E., Standardization of the one-stage prothrombin time for the control of anticoagulant therapy. Br Med J 1967; 1:84-90.

3. Box, G.E.P. (1979). Some problems of statistics and everyday life. Journal of the American Statistical Association, 74, 1-4.

4. Daniel, C. (1978). Patterns in residuals in the two-way layout. Technometrics, 20, 385-396.

5. Evatt, B.L., Brogan, D., Triplett, D., Koepke, J., Waters, G., Effect of thromboplastin and instrumentation on the prothrombin time test. Clin Lab Haematol 1981; 3:331-342.

6. Golomb, S.W. (1966). The information generating function of a probability distribution. IEEE Transactions on Information Theory, 12, 75-77.

7. Han, Y, Lawson, W.B., Dodds, W.J., and Lawrence, C.E., Prothrombin time proficiency testing, Thrombosis, and Haemostatis, 1985; 8:

8. Hermans, J, van den Besselaar, A.M.H.P., Loeliger, E.A., van der Velde, E.A., A collaborative study of reference materials for thromboplastins. Thromb Haemostas 1983; 50:712-717.

9. Ingram, G.I.C., Hills, M. Reference method for the one-stage prothrombin time test on human blood. Thromb Haemostas 1976; 36:237-238.

10. Kendall, M. and Stuart, A. The advanced theory of statistics 1976; 3:30.1. Loeliger, E.A., van Dijk-Wierda, C.A., van den Besselaar, A.M.H.P., Broekmans, A.W., Roos, J. Anticoagulant control and the risk of bleeding. In: Anticoagulants and myocardial infarction: A reappraisal. Meade, T.W. (Ed). John Wiley, New York 1984; 135-177.

11. Kullback, S. (1959). Information Theory and Statistics. New York: Wiley.

12. Loeliger, E.A., van Dijk-Wierda, C.A., Roos, J. Quality Control and standardization of the prothrombin time. Second International Symposium on Quality Control. Tokyo, Japan 1976.

13. Paulson, A.S., Delehanty, T.A., Sensitivity analysis in experimental design. In: Heiner, K.W., Sacher, R.S., Wilkinson, J.W. (Eds). Computer Science and Statistics: Proceedings of the 14th Symposium on the Interface. Springer-Verlag, New York 1982; 51-57.

14. Poller, L. The British system for anticoagulant control. Thromboe Diathes Haemorrh 1975; 33:32-50.

15. Queenberry, C.P. and David, H.A. (1961). Some tests for outliers. Biometrika, 48. 379-390.

16. Shapiro, S.S. and Wilk, M.B. (1965). An analysis of variance test for normality (complete samples). Biometrika, 52, 591-611.

17. Triplett, D.A., Evatt, B.L., van den Besselaar, A.M.H.P. Proficiency testing and standardization of prothrombin time - potential use of thromboplastion calibration in the United States. In: Ref 7, 209-221.

18. Van den Besselaar, A.M.H.P., Gralnick, H.R., Lewis, S.M., (Eds.). Thromboplastin Calibration and Oral Anticoagulant Control. Martinus Nijhoff, Boston, 1984.

19. van der Besselaar, A.M.H.P., Evatt, B.L., Brogan, D.R., Triplett, D.A., Proficiency testing and standardization of prothrombin time: Effect of thromboplastin instrumentation and plasma. Am J Clin Pathol 1984; 82:688-699.

MASTER COPY - FOR REPRODUCTION PURPOSES

UNCLASSIFIED

AD-A178 541

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1 REPORT NUMBER	2 GOVT ACCESSION NO. N A / 3 RECIPIENT'S CATALOG NUMBER N A
4 TITLE (and Subtitle)	5 TYPE OF REPORT & PERIOD COVERED
	6 PERFORMING ORG. REPORT NUMBER
7 AUTHOR(s)	8 CONTRACT OR GRANT NUMBER(s)
9 PERFORMING ORGANIZATION NAME AND ADDRESS	10 PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11 CONTROLLING OFFICE NAME AND ADDRESS	12 REPORT DATE
	13 NUMBER OF PAGES
14 MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15 SECURITY CLASS. (of this report)
	15a DECLASSIFICATION/DOWNGRADING SCHEDULE

16 DISTRIBUTION STATEMENT (of this Report)

17 DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

N A

18 SUPPLEMENTARY NOTES

19 KEY WORDS (Continue on reverse side if necessary and identify by block number)

20 ABSTRACT (Continue on reverse side if necessary and identify by block number)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

END

5-87

DTIC